

Implementasi Algoritma Rough Set Dan Naive Bayes Untuk Mendapatkan Rule Dalam Menyeleksi Pemohon Bantuan Fasilitas Rumah Ibadah (Studi Kasus : Pemerintah Kabupaten Pringsewu)

Jeprianto¹, RZ Abdul Aziz²

¹Program Studi Sistem Informasi, STMIK Pringsewu, Lampung

^{1,2}Program Pasca Sarjana Prodi Teknik Informatika IIB Darmajaya, Lampung, Indonesia

¹Jl. Wisma Rini No.09 Pringsewu, Lampung, Indonesia

^{1,2}Jl. ZA. Pagar Alam No.93, Gedong Meneng, Lampung, Indonesia

E-Mail : jevkiplingsing@gmail.com

Abstract-*In solving the problem of the accuracy of the selected algorithm if it is applied to a prototype application in predicting applicants for assistance with houses of worship facilities in Pringsewu District using data mining classification methods. In solving the problem using the rough set algorithm method and Naive Bayes from the results of the discussion carried out, it can draw conclusions Rough set algorithm and the resulting rule has the highest level of accuracy that is 92% Rough set algorithm model is included in the category of excellent classification and can be implemented in determining predictions more potential grant funding. The rules generated by the Rough set algorithm are applied in the prototype prediction of the grant of houses of worship grants with 92% accuracy of prototype verification testing results. Based on the accuracy of the resulting prototype shows that the methods and prototypes that are applied are good at predicting better results. Naive Bayes algorithm has an accuracy level of 77% The Naive Bayes algorithm model is included in the category of good classification and can be implemented in determining the prediction of grants but because the value of the rough set algorithm is higher then the naive Bayes algorithm is not used to determine the prototype.*

Keywords: *facilities, houses of worship, rough set and naive bayes algorithm methods, Pringsewu*

I. PENDAHULUAN

A. Latar Belakang Masalah

Seiring dengan berkembangnya teknologi informasi yang sangat pesat serta informasi yang

semakin meningkat, banyak sekali orang mengumpulkan informasi sebanyak mungkin menggunakan berbagai media yang dimiliki. Tetapi tidak semua orang mampu mengelola informasi dengan baik untuk dimanfaatkan pada waktu yang tepat dan secara efektif dan efisien. Memiliki banyak informasi saja tidak cukup jika tidak dapat mengelolanya menjadi alternatif terbaik dalam pengambilan keputusan. Teknologi informasi saat ini memungkinkan proses pengambilan keputusan dilakukan dengan cepat dan cermat.

Data mining merupakan bidang teknologi informasi yang berkembang pesat. Data mining melibatkan pemakaian database berskala besar maupun kecil. Informasi yang tersimpan dalam database menjadi tidak berguna seiring berjalannya waktu karena tidak dilakukan penggalian pengetahuan (*knowledge*) dari data-data yang banyak tersebut. Data mining dapat meningkatkan nilai tambah dari suatu database. Kita dapat menggali informasi yang tersimpan dalam database yang terakumulasi dalam jangka waktu lama untuk mendapatkan informasi tambahan.

Meningkatnya pengajuan proposal bantuan fasilitas rumah ibadah yang ditujukan kepada Pemerintah Kabupaten Pringsewu membuat pengambilan keputusan semakin sulit. Pemanfaatan aplikasi Microsoft Office Excel dalam pengolahan data penerima bantuan dana hibah membantu untuk menyeleksi pemohon bantuan berdasarkan dari nilai bobot setiap kriteria, hasil dari keputusan yang dikeluarkan berupa pemohon bantuan diterima, proses dan tolak. Berdasarkan dari keputusan yang dihasilkan maka tim seleksi harus mempertimbangkan dan meninjau kembali pemohon bantuan apakah layak diberikan bantuan atau tidak. Permasalahan yang

timbul adalah sulitnya dalam memberikan penilaian yang tepat dan adil disebabkan faktor jumlah data yang diproses relatif besar dan tidak adanya pola aturan yang tetap sehingga dalam pengambilan keputusan seringkali tidak jelas sehingga timbulnya perasaan buruk dari pemohon/calon penerima bantuan.

Agar memudahkan tim seleksi dalam menilai maka dibutuhkan pola aturan yang jelas agar menghasilkan keputusan yang tepat sasaran. penerapan data mining dapat membantu menggali informasi yang tersimpan dalam database dengan memanfaatkan Algoritma Rough Set dan Algoritma Naive Bayes. Dengan metode tersebut dapat menghasilkan informasi baru berupa pola aturan (rule) yang dapat digunakan dalam acuan penyeleksian pemohon bantuan fasilitas rumah ibadah, sehingga sangat membantu bagi tim seleksi dalam mengambil keputusan yang tepat sasaran.

B. Masalah Penelitian

Berdasarkan latar belakang masalah yang ada dalam aktivitasnya untuk mencapai tujuan yang diinginkan. Permasalahan yang dapat diidentifikasi pada penelitian ini adalah sebagai berikut:

1. Belum adanya sebuah metode klasifikasi prediksi dalam menentukan pola aturan pengajuan dana hibah rumah ibadah yang potensial.
2. Belum tersedianya sistem komputerisasi yg mampu memprediksi data-data pengajuan dana hibah rumah ibadah dengan lebih efektif dan efisien dengan menerapkan metode prediksi yang ada.

C. Tujuan dan Manfaat Penelitian

Dalam melakukan penelitian peneliti mempunyai maksud dan tujuan yang ingin dicapai, adalah sebagai berikut :

1. Meningkatkan akurasi prediksi dengan membandingkan metode terlebih dahulu untuk memprediksi pemohon bantuan fasilitas rumah ibadah di yang potensial .
2. Menghasilkan algoritma yang akurasinya lebih akurat untuk memprediksi pemohon bantuan fasilitas rumah ibadah dibandingkan penghitungan prediksi sebelumnya, sehingga penelitian selanjutnya dapat lebih meningkatkan lagi akurasinya.
3. Penelitian ini juga diharapkan dapat dijadikan bahan referensi terkait dengan pokok bahasan prediksi pemohon bantuan fasilitas rumah ibadah di Kabupaten Pringsewu.

Manfaat yang diperoleh dari hasil penelitian ini adalah sebagai berikut:

1. Meningkatkan akurasi prediksi dengan membandingkan 2 (dua) metode klasifikasi
2. Menghasilkan algoritma yang akurasinya lebih akurat untuk memprediksi data pemohon bantuan rumah ibadah, sehingga penelitian selanjutnya dapat lebih meningkatkan lagi akurasinya.
3. Instansi terkait dapat merencanakan perencanaan pemohon bantuan fasilitas rumah ibadah dengan lebih baik dan meningkatkan keakurasian aktual pemohon bantuan fasilitas rumah ibadah di Kabupaten Pringsewu.
4. Dengan meningkatkan keakurasian menggunakan salah satu metode klasifikasi maka pencapaian target bisa tercapai setiap harinya.

II. TINJAUAN PUSTAKA

A. Pengertian Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. (Turban, dkk. 2005)

Definisi umum dari *data mining* itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam *database*, *data warehouse*, atau media penyimpanan informasi yang lain. Hal penting yang terkait di dalam *data mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat. (Kusrini dan Emha Taufiq, 2009).

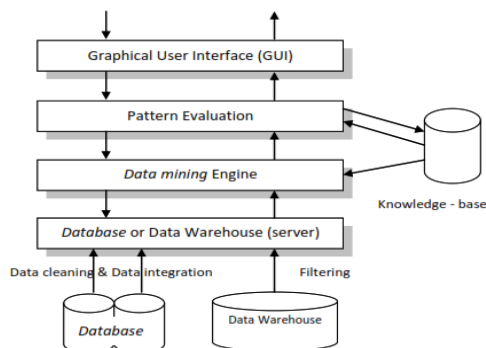
B. Arsitektur Sistem Data Mining

Arsitektur utama dari sistem *data mining*, pada umumnya terdiri dari beberapa komponen sebagai berikut:

1. *Database*, *data warehouse*, atau media penyimpanan informasi, terdiri dari satu atau beberapa *database*, *data warehouse*, atau data dalam bentuk lain. Pembersihan data dan integrasi data dilakukan terhadap data tersebut.
2. *Database*, *data warehouse*, bertanggung jawab terhadap pencarian data yang

relevan sesuai dengan yang diinginkan pengguna atau *user*.

3. Basis pengetahuan (*Knowledge Base*), merupakan basis pengetahuan yang digunakan sebagai panduan dalam pencarian pola.
4. *Data mining engine*, merupakan bagaian penting dari sistem dan idealnya terdiri dari kumpulan modul-modul fungsi yang digunakan dalam proses karakteristik (*characterization*), klasifikasi (*classification*), dan analisis kluster (*cluster analysis*). Dan merupakan bagian dari *software* yang menjalankan program berdasarkan algoritma yang ada.
5. Evaluasi pola (*pattern evaluation*), komponen ini pada umumnya berinteraksi dengan modul-modul *data mining*. Dan bagian dari *software* yang berfungsi untuk menemukan *pattern* atau pola-pola yang terdapat dalam *database* yang diolah sehingga nantinya proses *data mining* dapat menemukan *knowledge* yang sesuai.
6. Antar muka (*Graphical user interface*), merupakan modul komunikasi antara pengguna atau user dengan sistem yang memungkinkan pengguna berinteraksi dengan sistem untuk menentukan proses *data mining* itu sendiri.



Gambar 1. Arsitektur Data Mining

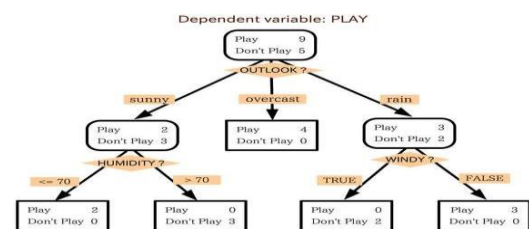
C. Teknik Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan basis data. Beberapa teknik yang sering disebut-sebut dalam literatur data mining antara lain yaitu :

Classification decision tree

Classification adalah suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan supervised induction, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas-kelas tambahan. Salah satu contoh yang mudah dan populer adalah dengan Decision tree, yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang prediksi kelas untuk contoh tersebut. Decision tree mudah untuk dikonversi ke aturan klasifikasi (classification rules).



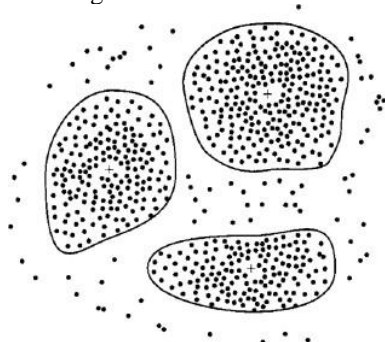
Gambar 2. Contoh classification decision tree

Association

Association digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana link asosiasi muncul pada setiap kejadian. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support yaitu prosentasi kombinasi atribut tersebut dalam basis data dan confidence yaitu kuatnya hubungan antar atribut dalam aturan asosiatif. Motivasi awal pencarian association rule berasal dari keinginan untuk menganalisa data transaksi supermarket, ditinjau dari perilaku customer dalam membeli produk. Association rule ini menjelaskan seberapa sering suatu produk dibeli secara bersamaan. Sebagai contoh, association rule "beer => diaper (80%)" menunjukkan bahwa empat dari lima customer yang membeli beer juga membeli diaper. Dalam suatu association rule $X \Rightarrow Y$, X disebut dengan antecedent dan Y disebut dengan consequent.

Clustering

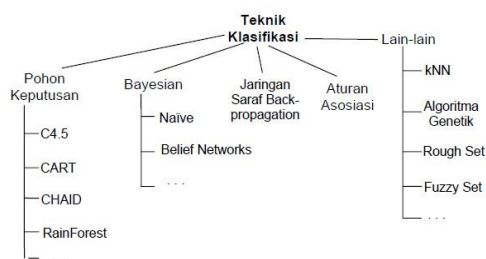
Clustering digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool data mining. Biasanya menggunakan metode neural network atau statistik. Clustering membagi item menjadi kelompok-kelompok yang ditemukan tool data mining.



Gambar 3. Contoh clustering

D. Klasifikasi naïve bayes

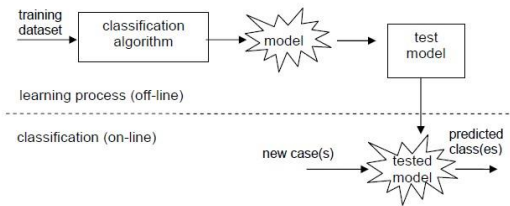
Teknik klasifikasi adalah suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data, dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Klasifikasi dalam data mining dikelompokkan ke dalam teknik pohon keputusan, Bayesian (Naïve Bayesian dan Bayesian Belief Networks), Jaringan Saraf Tiruan (Backpropagation), teknik yang berbasis konsep dari penambangan aturan-aturan asosiasi, dan teknik lain (k-Nearest Neighbor, algoritma genetik, teknik dengan pendekatan himpunan rough dan fuzzy). Setiap teknik memiliki kelebihan dan kekurangannya sendiri, berikut gambar pengelompokan teknik klasifikasi.



Gambar 4. Pengelompokan Teknik Klasifikasi

Secara umum, proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus baru. Pada proses belajar, algoritma klasifikasi mengolah data pelatihan untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan (Han et al.,2001; Quinlan, 1993).

Kelas yang dapat diprediksi adalah kelas-kelas yang sudah terdefinisi pada data pelatihan. Karena proses klasifikasi kasus baru cukup sederhana, penelitian lebih banyak ditujukan untuk memperbaiki teknik-teknik pada proses belajar.



Gambar 5. Skema Klasifikasi secara Umum

a. Algoritma Rough Set

Teori Rough set sampai saat ini pendekatan lain untuk ketidakjelasan (Pawlak, 1982). Demikian pula untuk teori himpunan fuzzy bukan merupakan alternatif untuk teorihimpunan klasik tetapi tertanam di dalamnya. Teori Rough Set dapat dilihat sebagai implementasi khusus dari gagasan G. Frege (1983) tentang ketidakjelasan, yaitu ketidaktepatan dalam pendekatan ini dinyatakan oleh batas wilayah dari suatu himpunan, dan bukan oleh keanggotaan parsial, seperti dalam teori himpunan fuzzy. Konsep Rough Set dapat didefinisikan cukup umum dengan cara operasi topologi, interior dan penutupan, yang disebut pendekatan.

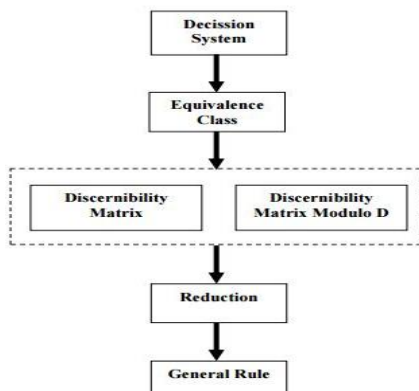
Tujuan analisis Rough Set adalah untuk mendapatkan rule yang klasifikasi setelah dilakukan pengumpulan data (Maharani, 2008). Rule disini sudah dikalsifikasikan setelah mendapatkan reduct.

Rough Set menentukan teorinya menggunakan perkiraan, yaitu yang ditentukan oleh fungsi keanggotaan. Rough Set bisa juga menentukan teorinya tanpa menggunakan perkiraan. Karena fungsi keanggotaan bukanlah konsep primitif dalam pendekatan yang dalam hal ini kedua defenisi tidak setara. (Jian, dkk 2011),

Fungsi keanggotaan merupakan pemetaan titik-titik yang didapat dari himpunan fuzzy kedalam keanggotaan yang memiliki interval antara 0 sampai dengan 1. Salah satu cara untuk mendapatkan nilai keanggotaan adalah dengan pendekatan fungsi. Di dalam Metode Rough Set terdapat beberapa langkah - langkah penyelesaian masalah, yaitu sebagai berikut:

1. Decision System tersebut dilakukan teknik klasifikasi kriteria yang disebut “Equivalen Class”
2. Kemudian dilakukan proses Discernibility Matrix atau Discernibility Matrix Modulo D
3. Proses “Reduction”
4. Untuk memperoleh hasil akhir dilakukan proses “General Rules” Langkah-langkah

dalam menjalankan metode rough set di atas dapat digambarkan pada gambar berikut ini.



Gambar 6. Proses Algoritma Rough Set

b. Algoritma Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Naive Bayes juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Saleh, 2015).

Persamaan dari teorema Bayes dapat dilihat di bawah ini :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots(2.1)$$

Dimana :

- X : data dengan class yang belum diketahui
- H : hipotesis data menggunakan suatu class spesifik
- P(H|X) : probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : probabilitas hipotesis H (prior probabilitas)
- P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : probabilitas H

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang di analisis tersebut. Karena itu, metode Naive Bayes di atas disesuaikan sebagai berikut (Saleh, 2015) :

$$P(C|F1 \dots Fn) = \frac{P(C) \cdot P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \dots\dots\dots(2.2)$$

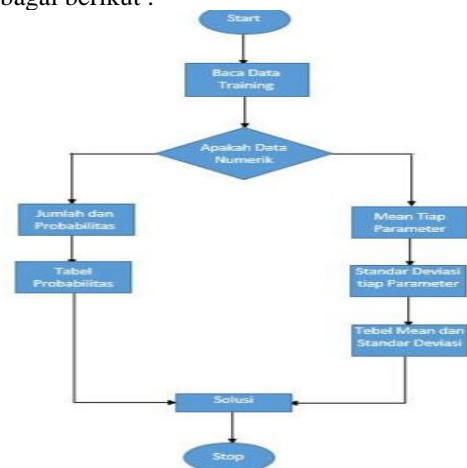
Di mana Variabel C mempresentasikan kelas, sementara variabel F1...Fn mempresentasikan karakteristik petunjuk yang dibutuhkan untuk menentukan klasifikasi. Maka rumus tersebut

menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut likelihood), dibagi dengan peluang kemunculan karakteristik – karakteristik secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut (Saleh, 2015):

$$posterior = \frac{prior \times likelihood}{evidence} \dots\dots\dots(2.3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari Posterior tersebut nantinya akan dibandingkan dengan nilai – nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Alur metode Naive Bayes dapat digambarkan sebagai berikut :



Gambar 7. Alur Metode Naive Bayes

Adapun persamaan yang digunakan untuk menghitung nilai rata – rata (mean) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots(2.4)$$

Atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \dots\dots\dots(2.5)$$

Di mana :

- μ : rata – rata hitung (mean)
- x_i : nilai sampel ke-i
- n : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots(2.6)$$

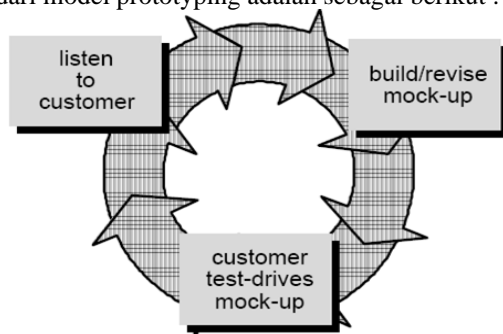
Dimana :

- σ : standar deviasi
- x_i : nilai x ke- i
- μ : rata – rata hitung
- n : jumlah sampel

- Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
- Mendapatkan nilai dalam tabel mean, standar deviasi dan probabilitas.
- Solusi yang dihasilkan

E. Prototyping

Salah satu metode pengembangan perangkat lunak yang banyak digunakan adalah *prototyping*. Selama proses pembuatan sistem, *developer* dan *client* dapat saling berinteraksi dengan menggunakan metode *prototyping* ini. Agar model *prototyping* berhasil adalah dengan mendefinisikan aturan-aturan yang harus disepakati *client* dan *developer* yaitu kesepakatan bahwa prototipe yang dibangun untuk mendefinisikan semua kebutuhan *client*. Prototipe akan dihilangkan sebagian atau seluruhnya dan perangkat lunak aktual direkayasa dengan kualitas dan implementasi yang sudah ditentukan. Bagan dari model *prototyping* adalah sebagai berikut :



Gambar 8. Model Prototyping
(Pressman, 2010)

Kelebihan *prototyping* adalah sebagai berikut:

- Adanya komunikasi yang baik antara *developer* dan *client*
- Developer* terfokus dengan kebutuhan *client*
- Client* berperan aktif dalam pengembangan sistem
- Pengembangan sistem membutuhkan waktu yang lebih efisien
- Penerapan menjadi lebih mudah, karena pengguna mengetahui apa yang diharapkan

Sementara itu, *prototyping* juga memiliki kelemahan yaitu:

- Client* terkadang kurang menyadari bahwa perangkat lunak yang ada belum mencantumkan kualitas perangkat lunak secara keseluruhan dan juga belum memikirkan kemampuan pemeliharaan untuk jangka waktu panjang.
- Developer* biasanya ingin proyek cepat selesai. Sehingga menggunakan

algoritma dan bahasa pemrograman sederhana agar prototipe lebih cepat selesai tanpa memikirkan bahwa *software* tersebut hanya merupakan *blue print* dari sistem yang akan dikembangkan kemudian.

- Hubungan *customer* dengan komputer yang disediakan mungkin tidak mencerminkan teknik perancangan yang baik

Prototyping bekerja dengan baik pada penerapan-penerapan yang berciri sebagai berikut (Sommerville, 2007):

- Resiko Tinggi, yaitu untuk masalah-masalah tidak terstruktur dengan baik, ada perubahan yang besar dari waktu ke waktu, dan adanya persyaratan data yang tidak menentu.
- Interaksi pemakai penting
Sistem harus menyediakan dialog *on-line* antara *costumer* dan komputer
- Perilaku pengguna yang sangat sulit ditebak atau mudah berubah
- Sistem yang inovatif. Sistem tersebut membutuhkan cara penyelesaian masalah dan penggunaan perangkat keras yang mutakhir
- Perkiraan tahap penggunaan ssstem yang pendek.

III. METODOLOGI PENELITIAN

A. Metode Penelitian

Menurut (Dawson 2009, 87) ada empat metode penelitian yang umum digunakan yaitu tindakanpenelitian, eksperimen, studi kasus dan *survey*. Dalam konteks penelitian ini menggunakan eksperimen, yaitu suatu metode yang dilakukan dengan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis, pengujian hipotesis, menafsirkan hasil, dan kesimpulan (Berndtssom 2008, 27).

Penelitian ini jika dilihat dari bentuk data dan informasi yang dikelola, penelitian ini tergolong jenis penelitian kuantitatif. Penelitian kuantitatif adalah penelitian yang hipotesisnya dapat diuji dengan teknik-teknik statistik (Moedjiono 2016).Metode ini digunakan saat melakukan pengujian kualitas yaitu menggunakan metode *k-fold cross validation* ($k=10$) yang menampilkan nilai akurasi, *precision*, *recall*, ROC pada masing-masing metode yang dibandingkan.

B. Metode Pemilihan Sampel

Teknik sampling adalah teknik yang dilakukan dalam pengambilan sampel yang akan digunakan dalam penelitian. Metode pemilihan sample yang digunakan adalah sampling jenuh. Sampling jenuh adalah teknik penentuan sampel bila semua anggota populasi digunakan sebagai sampel. Hal ini sering dilakukan bila jumlah populasi relatif kecil, kurang dari 30 orang, atau penelitian yang ingin membuat generalisasi dengan kesalahan yang sangat kecil. Istilah lain sampel jenuh adalah sensus, dimana semua anggota populasi dijadikan sampel. (Sugiyono, 2001)

C. Tahap Pengumpulan Data

Dalam penelitian ini metode yang digunakan dalam proses pengumpulan data antara lain:

a. Studi Pustaka

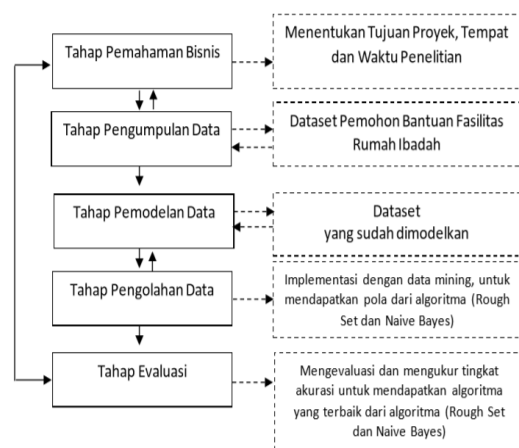
Studi pustaka dilakukan dengan cara menghimpun informasi yang relevan dengan topik atau masalah yang akan atau sedang diteliti. Informasi diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, ensiklopedia, dan sumber-sumber tertulis baik tercetak maupun elektronik.

b. Dokumentasi

Pengambilan data dilakukan melalui dokumen tertulis maupun elektronik dari lembaga/institusi. Dokumen diperlukan untuk mendukung kelengkapan data yang lain.

c. Seleksi Data

Sebelum memasuki tahap transformasi data dilakukan, maka perlu dilakukan seleksi data, dari beberapa variabel data yang ada, variabel yang dipilih untuk digunakan dalam tahap pengolahan data antara lain: Kelengkapan Berkas, Tahun Pendirian Rumah Ibadah, Kondisi Rumah Ibadah dan Tingkat Rutinitas Rumah Ibadah. Penelitian ini dilaksanakan berdasarkan alur penelitian seperti yang ditunjukkan pada Gambar 9.



Gambar 9. Alur Penelitian

IV. HASIL DAN PEMBAHASAN

A. Pengelompokan dan analisis data

Data mining tidak akan pernah dapat dilakukan tanpa tersedianya data yang akan di tambang. Untuk itu dibutuhkan pemahaman akan data yang seperti apa yang akan dipakai dalam proses data mining nanti. Dalam bagian ini akan dibahas mengenai pengolahan data yang diperoleh dari hasil penelitian dilapangan. Data yang digunakan dalam penelitian ini yaitu: Data *Decision System (DS)* yaitu berupa daftar hasil seleksi pemohon bantuan dana hibah fasilitas rumah ibadah tahun 2019 sebanyak 100 record.

Data yang dipakai dalam proses data *mining*, perlu dipersiapkan dan diubah dalam bentuk model yang lebih sederhana agar bisa dianalisis dengan memakai teknik data mining. Menurut Peraturan Bersama Menteri Agama Dan Menteri Dalam Negeri Nomor: 9 Tahun 2006 dan Nomor: 8 Tahun 2017. Adapun beberapa jenis dan kriteria yang menjadi acuan untuk menseleksi pemohon bantuan dana hibah fasilitas rumah ibadah, antara lain seperti tabel dibawah ini. Data yang dipakai secara umum dibagi dalam 2 (dua) jenis data, yaitu Verifikasi data Monitoring.

Verifikasi data adalah data yang wajib dipenuhi, jika tidak maka proposal tidak akan diproses ketahap monitoring. Berikut adalah tabel jenis dan kriteria penilaian.

Tabel 1. Kriteria pemilihan variable

No.	Nama Variabel	Komponen Penilaian	Nilai	Transformasi Data
1	Kelengkapan Berkas	Verifikasi Berkas dan Proposal	√	Ya
			X	Tidak
2	Tahun Pendirian	Surat izin Tahun pendirian	≥2010	Baru
			< 2010	Lama
3	Kondisi Rumah	Kondisi Bangunan Luar	≤ 4	Kurang
			≤ 6	Cukup
			≤ 8	Baik
		Kondisi Bangunan Dalam	≤ 4	Kurang
			≤ 6	Cukup
			≤ 8	Baik
		Kondisi bangunan Pendukung	≤ 4	Kurang
			≤ 6	Cukup
			≤ 8	Baik
4	Antusias masyarakat sekitar	≤ 4	Kurang	
		≤ 6	Cukup	
		≤ 8	Baik	
5	Rutinitas Kegiatan Rohani	≤ 4	Kurang	
		≤ 6	Cukup	
		≤ 8	Baik	

Dari kelima kriteria pada tabel 4.1 diantaranya memiliki nilai bobot berupa persentase sebagai acuan dalam menyeleksi pemohon bantuan.

B. Algoritma Metode Rough Set

Karena data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu dilakukan proses seleksi data (*Data Selection*), maka didapatkan data yang akan kemudian diolah berjumlah 100 pemohon dimana data dikelompokkan dalam 19 kelompok yang sama.

Setelah data didapat dan dirangkap maka tahapan pertama yaitu menentukan atribut

keputusan berdasarkan data monitoring, atribut keputusan yang akan dibuat adalah terima, tolak atau proses yang artinya Terlihat pada tabel 4.3.

Tabel 4.3 Variabel Keputusan

Output	Terima
	Proses
	Tolak

Langkah selanjutnya adalah memberikan symbol pada semua atribut yang ada. Simbol-simbol semua atribut dapat dilihat pada tabel 4.4

Tabel 4.4 Simbol variabel

No	Variabel	Simbol
1.	Persyaratan	A
2.	Tahun Pendirian	B
3.	Kondisi Bangunan	C
4.	Antusias Masyarakat	D
5.	Rutinitas tempat ibadah	E

Equivalence class

Equivalence Class merupakan pengelompokan objek-objek yang sama berdasarkan atribut tertentu. Pembentukan *Equivalence Class* dilakukan dengan cara mengelompokkan data *Decision System (DS)* yang memiliki kesamaan kedalam satu *class*, adapun hasil dari pembentukan *Equivalence Class* dapat dilihat pada table 2

Tabel 2 Equivalence class

CLASS	P	TP	KB	A	R	K	JUMLAH
	A	B	C	D	E	K	
EC1	YA	LAMA	KURANG	KURANG	CUKUP	PROSES	9
EC2	YA	LAMA	KURANG	KURANG	KURANG	TERIMA	3
EC3	YA	LAMA	KURANG	CUKUP	KURANG	PROSES	4
EC4	YA	LAMA	KURANG	CUKUP	CUKUP	TERIMA	4
EC5	YA	LAMA	CUKUP	KURANG	CUKUP	PROSES	7
EC6	YA	LAMA	CUKUP	BAIK	CUKUP	TERIMA	2
EC7	YA	LAMA	BAIK	KURANG	BAIK	TERIMA	4
EC8	YA	LAMA	BAIK	BAIK	CUKUP	TOLAK	2
EC9	YA	LAMA	BAIK	BAIK	BAIK	TOLAK	5
EC10	YA	BARU	KURANG	KURANG	CUKUP	PROSES	5
EC11	YA	BARU	KURANG	CUKUP	CUKUP	TERIMA	10
EC12	YA	BARU	KURANG	CUKUP	KURANG	TOLAK	5
EC13	YA	BARU	CUKUP	KURANG	CUKUP	PROSES	8
EC14	YA	BARU	CUKUP	CUKUP	KURANG	TOLAK	8
EC15	YA	BARU	CUKUP	CUKUP	CUKUP	TERIMA	6
EC16	YA	BARU	BAIK	CUKUP	CUKUP	TOLAK	3
EC17	YA	BARU	BAIK	KURANG	BAIK	TOLAK	7
EC18	YA	BARU	BAIK	CUKUP	BAIK	PROSES	3
EC19	YA	BARU	BAIK	BAIK	BAIK	TOLAK	5
TOTAL							100

Dari total 100 data didapatkan 19 class untuk data uji. Berdasarkan hasil penyederhanaan *aljabar Boolean* secara manual didapatkan *reduct* seperti tabel diatas bahwa *reduct* yang dihasilkan terdiri dari kombinasi atribut sebagai berikut :

1. {B} = Tahun Pendirian (TP)
2. {C}= Kondisi (K)
3. {D}= Antusias Masyarakat (A)

4. {E} = Rutinitas (R)
5. {B,E} = Tahun Pendirian (TP) dan Rutinitas (R)
6. {B,D} = Tahun Pendirian (TP) dan Antusias Masyarakat (A)
7. {C,D} = Kondisi (K) dan Antusias Masyarakat (A)
8. {C,E} = Kondisi (K) dan Rutinitas (R)
9. {D,E} = Antusias Masyarakat (A) dan Rutinitas (R)
10. {B,C,D} = Tahun Pendirian (TP), Kondisi (K), dan Antusias Masyarakat (A).
11. {B,D,E} = Tahun Pendirian (TP), Antusias Masyarakat (A), dan Rutinitas (R).
12. {C,D,E} = Kondisi (K), Antusias Masyarakat (A), dan Rutinitas (R).

C. Evaluasi dan Validasi

Hasil dari pengujian model yang telah dilakukan yaitu dengan algoritma Naïve bayes, dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix*. Dari hasil *confusion matrix*, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, dan *recall* yang telah dihitung untuk metode Naïve bayes dan Rough set dapat dilihat pada Tabel 4.9.

Tabel 3. Perbandingan Nilai *Accuracy*, *Precision*, dan *Recall*

	Naïve Bayes	Rough set
	Training	Training
Accuracy	77%	92%
Precision	85,48%	96,67%
Recall	91,38%	100%

Berdasarkan pengujian dan evaluasi hasil klasifikasi dengan algoritma Naïve bayes dan *rough set*, dapat kita lihat hasilnya yaitu tingkat akurasi yang paling tinggi dengan algoritma Rough set dengan tingkat akurasi 92%.

Dengan menggunakan perbandingan data *training*, maka perbandingan komparasinya untuk akurasi dapat dilihat dalam tabel 4.10:

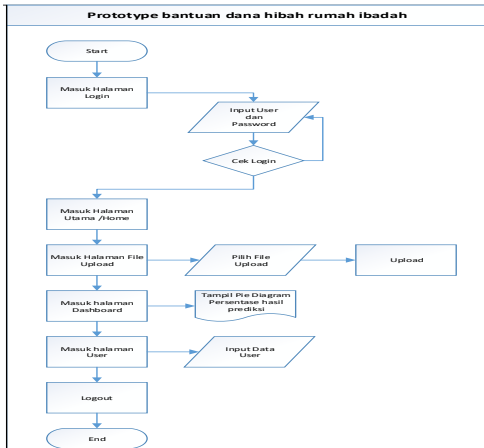
Tabel 4. Perbandingan Akurasi

	Naïve Bayes	Rough set
Accuracy	77%	92%

Berdasarkan hasil perbandingan akurasi pada tabel 4.10 Algoritma Rough set yang memiliki tingkat akurasi yang paling tinggi, sehingga baik digunakan untuk prediksi pemberian dana hibah rumah ibadah dengan persentase 92%.

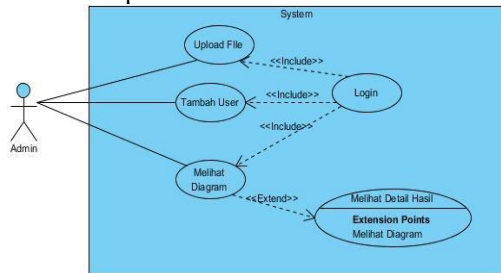
D. *Prototype Algoritma Terpilih*

Dari hasil evaluasi dan validasi diatas dapat diketahui bahwa Algoritma *naive bayes* memiliki tingkat akurasi dan performansi yang baik, sehingga *rule* yang dihasilkan oleh algoritma *Rough set* dapat dijadikan sebagai *rule* untuk pembuatan *prototype* yang dapat memudahkan dalam memprediksi keputusan penerimaan bantuan dana hibah rumah ibadah.



Gambar 10. Flow Chart Prototype bantuan dana hibah rumah ibadah

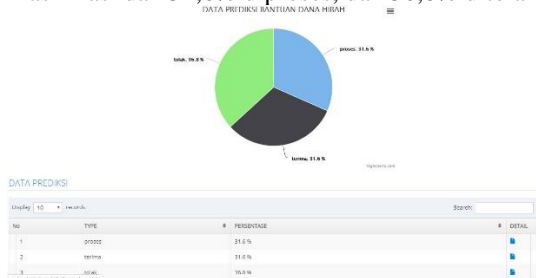
Uses case aplikasi



Gambar 11. Use case aplikasi

E. Hasil Prediksi

Hasil penerapan *rule* algoritma *Rough set* terhadap data baru sejumlah 19 record data dimana 31,6% data diprediksi *diterima* sesuai dengan hasil klasifikasi dan 31,6% diproses, dan 36,8% ditolak.



Gambar 12. Form Dashboard

Dari hasil verifikasi algoritma terpilih menggunakan aplikasi berbasis web dari 19 record didapatkan penerima dana hibah yang terprediksi 6 rumah ibadah dengan presentase sebesar 31,6%.

No.	Nama Rumah Ibadah	Alamat	Tanggal Pendirian	Tahun Pendirian	Kondisi Bangunan Luar	Antusias Masyarakat	Kondisi Bangunan Dalam	Persyaratan	Rutinitas Pemakaian
1	AL MUSAFAWA	KEL PSW SELATAN	18/08/2018	baru	cukup	cukup	cukup	ya	cukup
2	AL AMIN	RT.01 RW.02 PAJARESUK	30/01/2017	baru	kurang	cukup	kurang	ya	cukup
3	HIDAYATUSS	RT.04 RW.04 PSW TIMUR	10/01/2005	lama	baik	kurang	baik	ya	baik
4	MUFIDUDDIN	RT.03 RW.03 PSW TIMUR	06/08/2009	lama	cukup	baik	cukup	ya	cukup
5	AL HIDAYAH	RT.03 RW.03 PSW TIMUR	28/01/2005	lama	kurang	cukup	kurang	ya	cukup
6	NURUL IMAN	PEKON FAJAR AGUNG	09/09/2007	lama	kurang	kurang	kurang	ya	kurang

Gambar 12. Form Details Terima

Dari hasil verifikasi algoritma terpilih menggunakan aplikasi berbasis web dari 19 record didapatkan penerima dana hibah yang terprediksi akan Proses sebanyak 6 pemohon, dengan presentase sebesar 31,6%.

No.	Nama Rumah Ibadah	Alamat	Tanggal Pendirian	Tahun Pendirian	Kondisi Bangunan Luar	Antusias Masyarakat	Kondisi Bangunan Dalam	Persyaratan	Rutinitas Pemakaian
1	AL HIDAYAH	KEL PSW SELATAN	23/02/2012	baru	baik	cukup	baik	ya	baik
2	BATUL MAH	RT.01 RW.04 PAJARESUK	18/08/2017	baru	cukup	kurang	cukup	ya	cukup
3	NURUL HUDA	RT.05 RW.01 PAJARESUK	13/08/2012	baru	kurang	kurang	kurang	ya	cukup
4	AL WUSTHO	RT.01 RW.01 PSW TIMUR	28/02/2005	lama	cukup	kurang	cukup	ya	cukup
5	AL HIDAYAH	PEKON FAJAR AGUNG	10/02/2008	lama	kurang	cukup	kurang	ya	kurang
6	AL WILAS	PEKON FAJAR AGUNG	04/04/2007	lama	kurang	kurang	kurang	ya	cukup

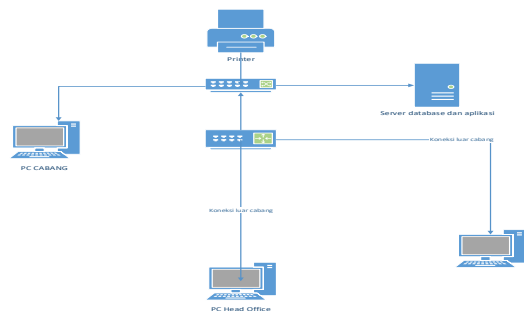
Gambar 13. Form Details Proses

Dari hasil verifikasi algoritma terpilih menggunakan aplikasi berbasis web dari 19 record didapatkan penerima dana hibah yang terprediksi akan Tolak sebanyak 7 orang, dengan presentase sebesar 36,8%.

No.	Nama Rumah Ibadah	Alamat	Tanggal Pendirian	Tahun Pendirian	Kondisi Bangunan Luar	Antusias Masyarakat	Kondisi Bangunan Dalam	Persyaratan	Rutinitas Pemakaian
1	AL MUHAJIR	KEL PSW SELATAN	10/03/2002	baru	baik	baik	baik	ya	baik
2	AL MUTAQIM	KEL PSW SELATAN	27/12/2005	baru	baik	kurang	baik	ya	baik
3	SUHADA	KEL PSW SELATAN	17/11/2007	baru	baik	cukup	baik	ya	cukup
4	AL FAJRI	RT.02 RW.03 PAJARESUK	14/04/2008	baru	cukup	cukup	cukup	ya	kurang
5	BADRUL MUN	RT.01 RW.04 PAJARESUK	01/01/2007	baru	kurang	cukup	kurang	ya	kurang
6	NURROHMAD	RT.01 RW.01 PAJARESUK	17/04/2008	lama	baik	baik	baik	ya	baik
7	AL MUHAJIR	RT.01 RW.05 PSW TIMUR	18/02/2008	lama	baik	baik	baik	ya	cukup

Gambar 14. Form Details Tolak

Berikut ini desain arsitektur sistem data mining:



Gambar 4.14 Arsitektur device datamining

F. *Pengujian Fungsional Prototype dengan Blackbox*

Pengujian dilakukan dengan tujuan untuk mengetahui apakah aplikasi dibangun sesuai dengan fungsional yang di harapkan.

Tabel 5. Pengujian Kotak Hitam

Kelas Uji	Butir Uji	Jenis Pengujian
File Upload	Pilih File	Black Box
	Upload File	Black Box
Dashboard	Lihat Grafik	Black Box
User	list user	Black Box
	Tambah user	Black Box
	Delete user	Black Box

V. SIMPULAN

Dari pengukuran kinerja dengan melakukan komparasi dua algoritma yang telah dilakukan berdasarkan jumlah data maka dapat disimpulkan bahwa algoritma *Rough set* memiliki kemampuan dalam pengambilan keputusan untuk menentukan keputusan bantuan dana hibah rumah ibadah. Kedua algoritma *Naïve bayes dan rough set* dapat digunakan dalam menentukan keputusan bantuan dana hibah rumah ibadah. Kedua algoritma ini dikomparasi kemudian diuji akurasi. Tingkat akurasi tertinggi lah yang digunakan dalam menentukan rumah ibadah potensial penerima dana hibah.

1. Algoritma *Rough set* beserta rule yang dihasilkan memiliki tingkat akurasi tertinggi yaitu 92% Model algoritma *Rough set* termasuk dalam kategori *excellent classification* dan bisa di implementasikan dalam menentukan prediksi pemberian dana hibah yang lebih potensial.
2. *Rule* yg dihasilkan algoritma *Rough set* diterapkan dalam *prototype* prediksi pemberian dana hibah bantuan rumah ibadah dengan hasil akurasi pengujian verifikasi *prototype* sebesar 92%. Berdasarkan akurasi yg dihasilkan *prototype* menunjukkan bahwa metode dan *prototype* yg diterapkan sudah baik dalam memprediksi hasil yang lebih baik.
3. Algoritma *Naïve Bayes* memiliki tingkat akurasi yaitu 77% Model algoritma *Naïve bayes* termasuk dalam kategori *good classification* dan bisa di implementasikan dalam menentukan prediksi pemberian dana hibah tetapi karena nilai algoritma *rough set* lebih tinggi maka algoritma *naïve bayes* tidak dipakai untuk penentuan pembuatan *prototype*.

DAFTAR PUSTAKA

- [1] Bramer, Max. (2007). *Principles of Data Mining*. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7.
- [2] Berndtsson 2008 Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. *A Guide For Students In Computer Science And Information Systems*. London: Springer, 2008.
- [3] Dawson 2009 Dawson, C. W. *Projects In Computing And Information System A Student's Guide*. England: Addison-Wesley, 2009.
- [4] Elvira Sukma Wahyuni. 2015. *Penerapan Metode Seleksi Fitur Untuk Meningkatkan Hasil Diagnosis Kanker Payudara*. Jurnal teknologi
- [5] Gorunescu, F. (2011). *Data Mining Concept Model and Techniques*. Berlin: Springer. ISBN 978-3-642-19720-8
- [6] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- [7] Kotler 2009, 245 Batubara, Muhamad Hendri. *Strategi Marketing Public Relation (MPR) Berupa Promosi Dan Sponsorsip Untuk Mempengaruhi Konsumen Dalam Keputusan Pembelian (Studi Produk Perawatan Bayi Johnsons Baby)*. Tes., Universitas Indonesia, 2010.
- [8] Kusriani 2009 Kusriani. *Algoritma Data Mining*. Jakarta: Andi, 2009.
- [9] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- [10] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc. ISBN 0-471-66657-2.
- [11] Liao, T. W., & Triantaphyllou, E. (2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications. Series on Computers and Operations Research, Vol 6. USA: World Scientific*. ISBN-13 978-981-277-985-4, ISBN-10 981-277-985-X.
- [12] Ulfah Indriani. 2018. *Penerapan Metode Rough Set Dalam Menentukan Pembelian Smartphone Android Oleh Konsumen*. Jurnal teknologi